Dispersion in Dispersion: Measuring Establishment-Level Differences in Productivity

by

Cindy Cunningham
Bureau of Labor Statistics

Lucia Foster
U.S. Census Bureau

Cheryl Grim
U.S. Census Bureau

John Haltiwanger
University of Maryland

Sabrina Wulff Pabilonia
Bureau of Labor Statistics

Jay Stewart
Bureau of Labor Statistics

Zoltan Wolf
Westat

CES 18-25   April, 2018

Abstract

Productivity measures are critical for understanding economic performance. The official Bureau of Labor Statistics (BLS) productivity statistics, which are available for major sectors and detailed industries, are useful for understanding the sources of aggregate productivity growth. A large volume of research shows that within-industry variation in productivity provides important insights into productivity dynamics. This research has revealed large and persistent productivity differences across businesses even within narrowly-defined industries. These differences vary across industries and over time and are related to productivity-enhancing reallocation. Dispersion in productivity across businesses may provide information about the nature of competition and frictions within sectors and about the sources of rising wage inequality across businesses. There are currently no official statistics that provide this level of detail. To fill this gap in the official statistics, the BLS and the Census Bureau are collaborating to create measures of within-industry productivity dispersion with the goal of publishing dispersion statistics to complement the official aggregate and industry-level productivity growth statistics produced by the BLS and thereby improve our understanding of the rich productivity dynamics in the U.S. economy. We are also developing restricted-use datasets for use by researchers in the Federal Statistical Research Data Center
(FSRDC) network.

**Keyword*:*

JEL Classification:

1. Introduction

Productivity measures are critical for understanding economic performance in the U.S. economy. The Bureau of Labor Statistics (BLS) produces the official labor and multi-factor productivity statistics for major sectors and industries in the U.S. These statistics are constructed using aggregate industry-level data, and can be thought of as changes in the first moment of establishment-level productivity (appropriately weighted). That is, these statistics show how productivity changes *on average* within sectors and industries, but they cannot provide insight into the variation in productivity levels across establishments within sectors or industries.[1]

To fill this void, the BLS and the Census Bureau initiated the Collaborative Micro-productivity Project (CMP) to develop and publish official statistics on *within-industry* productivity dispersion (i.e., second-moment measures of establishment-level productivity) and to produce restricted-use datasets. The public-use statistics would cover industries in the manufacturing sector and would be published jointly by the BLS and the Census Bureau. Restricted-use establishment-level data would also be made available to qualified researchers on approved projects in secure Federal Statistical Research Data Centers (FSRDCs).[2]

Economic theory and recent empirical evidence suggest the second moments of productivity are informative on a number of important dimensions. One of the most important findings in the literature on micro-level productivity is that there are large productivity differences across establishments even within narrowly-defined industries.[3] For example, using data from the 1977 Census of Manufactures (CM), Syverson (2004b) found that establishments at the 90th percentile of the within-4-digit-SIC productivity distribution are about four times as productive as

---

[1] Although usually referred to as industry productivity growth or aggregate productivity growth, these statistics can be thought of as the weighted average of within-industry growth rates.

[2] For more information on the FSRDCs, see the FSRDC website: http://www.census.gov/fsrdc. An earlier version of this dataset was analyzed in Foster et al. (2016a).

[3] Syverson (2011) provides a survey of this literature.

1

those at the 10th percentile.

Syverson's findings have generated considerable interest in the causes and consequences of such dispersion. Possible market explanations include curvature of the profit function that prevents the most-productive business from taking over an industry, frictions in factor adjustments (such as costs of adjusting input factors), barriers to entry and exit, and distortions that inhibit the equalization of marginal products across businesses (such as the regulatory environment). Drivers of establishment-level productivity variation include differences in management skills, the quality of production factors, innovation, and investments in R&D.

Research has shown that the dispersion of establishment-level productivity varies across sectors, by geographic area, and over time. For example, Syverson (2004a, 2004b) shows that variation in dispersion across industries and geographic areas is related to product substitutability, market structure, and competition. Hsieh and Klenow (2009) argue that both cross-country variation and within-country variation in the dispersion of productivity are related to distortions that inhibit productivity-enhancing reallocation. Asker, Collard-Wexler, and De Loecker (2014) present evidence that the patterns of dispersion reflect the dynamic factor adjustment frictions within sectors. The findings in Foster et al. (2016b) suggest that productivity differences across establishments may be generated by differences in efficiency levels, demand shocks, frictions/distortions or all of the above. Alternatively, Foster et al. (2017b) present evidence that industries experiencing a surge in innovation exhibit a burst of firm entry, followed by an increase in productivity dispersion during an experimentation and shakeout phase, followed ultimately by an increase in productivity.

Establishment-level productivity differences are also correlated with important economic outcomes at the micro level, such as the survival and growth of establishments. There is a large literature on the connection between productivity, reallocation and growth (Griliches and Regev, 1992; Baily, Hulten, and Campbell, 1992; Baily, Bartelsman, and Haltiwanger, 2001; Petrin,

White, and Reiter, 2011; Hsieh and Klenow, 2009, 2014). These studies show that more-productive businesses are more likely to survive and grow. This implies that reallocation—the process by which low-productivity businesses contract and exit while high-productivity businesses survive and expand—is an important contributor to aggregate productivity growth.

Productivity dispersion is also important for understanding rising wage inequality, which has been shown to be a between-firm phenomenon (Davis and Haltiwanger, 1991; Bloom et al., 2015; Barth et al., 2016). In addition, a number of studies have found that high-wage establishments are also highly-productive and that rising between-establishment dispersion in wages is closely associated with rising between-establishment dispersion in productivity (e.g., Dunne et al., 2004). Economic theories of search and matching provide theoretical justification for the connection between productivity dispersion and wage dispersion (e.g., Burdett and Mortensen, 1998). Search and matching frictions create quasi-rents for worker-firm matches that make it optimal for high-productivity firms to pay high wages.

Because one objective of the CMP is to publish public-use statistics on productivity dispersion to complement the official BLS data, it is crucial to understand the relationship between the dispersion of the productivity distribution derived from Census Bureau microdata and the statistics from BLS built from industry-level aggregates. Section 2 describes BLS productivity measures and productivity measures we construct from Census microdata. Section 3 compares the two approaches to measuring inputs, output, and labor productivity measures for the manufacturing sector, and for 4-digit manufacturing industries.[4] We also compare these measures to data from the NBER-CES Manufacturing Industry Database, and examine a number of data and measurement issues such as imputation and weighting of the microdata. In section 4, we explore the variation in our industry-level labor productivity dispersion measures across industries and

---

[4] In the next draft of this paper, we will also construct establishment-level multifactor productivity (MFP) measures. In general terms, MFP is defined as output per unit of an input index where the latter is a weighted average of input factors. Assuming a Cobb-Douglas production function, the weights are the factor elasticities.

over time. Section 5 summarizes our conclusions and describes plans for future work.

2. Measuring Productivity

Since a primary goal of this project is to create within-industry productivity dispersion statistics to complement BLS industry productivity data, it is useful to first describe how the BLS constructs its measures from published aggregates, and then compare it to our measures that are constructed by aggregating Census microdata.

2.1. BLS Industry-level Productivity

The BLS publishes quarterly and annual measures of labor productivity for major sectors; annual measures of labor productivity for 199 3-digit and 4-digit NAICS industries; and annual measures of multifactor productivity for major sectors, 18 3-digit NAICS manufacturing industries, 86 4-digit NAICS manufacturing industries, the air transportation industry, and the line-haul railroad industry. Productivity growth is measured as the difference between percentage changes in output and inputs (calculated as indexes). The BLS does not publish industry productivity *levels*, although they are available on request.

BLS industry output is based on a sectoral concept, which measures the value of goods produced for sale outside the industry.[5] For manufacturing industries, the BLS uses published Annual Survey of Manufactures (ASM) and Census of Manufactures (CM) data on the total value of shipments, which it adjusts to remove intrasectoral transactions and resales, and to account for changes in finished goods and work-in-process inventories.[6] This adjusted nominal output measure is then distributed to detailed categories of products and services. Nominal output in each

---

[5] Sectoral output is less than gross output, but greater than value-added output. In the most detailed industries, sectoral and gross output are the same or very close. However, going from very detailed industries to more aggregated industries, sectoral output moves closer to value-added output. In the limit, at the aggregate level, sectoral output is the same as value-added output, except for imported intermediate inputs.
[6] See https://www.census.gov/programs-surveys/asm.html and https://www.census.gov/econ/www/mancen.html.

category is deflated using the appropriate detailed producer price index from the BLS prices program. These real output measures are then Tornqvist-aggregated into industry output indexes. Self-employment revenues for manufacturing firms, which come from Internal Revenue Service data, are also added to these output measures.

The BLS measures labor input as the total annual hours worked by all persons in an industry. This measure is constructed by combining data from three BLS surveys: the Current Employment Statistics (CES) survey, the Current Population Survey (CPS), and the National Compensation Survey (NCS). The CES provides detailed information on the employment and average weekly hours *paid* for production and non-supervisory employees (henceforth referred to as production workers).[7] The NCS data are used to adjust CES hours from an hours-paid to an hours-worked basis by removing paid vacation accrued and sick leave taken.[8] To estimate nonproduction worker hours, the BLS uses data from the CPS to calculate the ratio of nonproduction to production worker hours worked, which is then multiplied by the adjusted CES production worker hours (worked). Specifically, nonproduction worker hours are estimated as:

$$\text{TH}_{NP} = \text{Emp}_{\text{NP}}^{\text{CES}} \times \text{AWH}_{\text{P}}^{\text{CES}} \times hwhp_P^{NCS} \times \frac{\text{AWH}_{\text{NP}}^{\text{CPS}}}{\text{AWH}_{\text{P}}^{\text{CPS}}} \times 52 \tag{1}$$

where $\text{Emp}_{\text{NP}}^{\text{CES}}$ is nonproduction worker employment from CES, $\text{AWH}_{\text{P}}^{\text{CES}}$ is production worker average weekly hours from CES, $hwhp_P^{NCS}$ is the hours-worked-to-hours-paid ratio from NCS, and $\left(\frac{\text{AWH}_{\text{NP}}^{\text{CPS}}}{\text{AWH}_{\text{P}}^{\text{CPS}}}\right)$ is the CPS nonproduction/production hours ratio. CPS data are also used to directly obtain hours worked by self-employed and unpaid family workers (Eldridge et al., 2004).

---

[7] Workers in goods-producing industries are referred to as being production or non-production workers and in the service-providing industries as nonsupervisory or supervisory workers.

[8] Note that this adjustment does not account for off-the-clock hours.

2.2. Establishment-level Productivity using Census Data

To measure establishment-level labor productivity, we combine establishment-level information from three Census Bureau restricted-use microdata files with public-use industry-level data from BLS. Given that one of the goals of our research is to shed light on BLS industry productivity statistics, we try to match BLS concepts and measures as closely as possible.

Our establishment-level microdata come from the CM, the ASM, and the Longitudinal Business Database (LBD). The CM is collected every 5 years in years ending in '2' and '7'. It collects data from all manufacturing establishments except those that are very small. The Census Bureau imputes data for these very small establishments using information from administrative records. The ASM sample is a 5-year panel of manufacturing establishments, updated every year for births, and data are collected annually. ASM panels begin in years ending in '4' and '9', and the probability of selection into the ASM sample is a function of both industry and size (generally employment or shipments). Like the CM, the ASM does not collect data from very small establishments but accounts for them using administrative information. In CM years, ASM data are collected as part of the CM data collection, but we use only the ASM establishments to create our within-industry dispersion measures.[9] Data are imputed for establishments that do not respond or that fail to report some data elements (item non-response); we discuss this further in section 2.3. The LBD is a longitudinally-linked version of the Census Bureau's Business Register and covers the non-agricultural employer universe of business establishments (see Jarmin and Miranda, 2002). The LBD provides us with both high-quality longitudinal links and information on the universe of manufacturing establishments, which we use to generate our propensity-score weights..

We cannot exactly replicate the BLS sectoral output concept because the ASM does not collect information on who the output was sold to or which industries intermediate inputs were

---

[9] The microdata made available in the FSRDCs will contain productivity measures for all CM establishments when productivity calculation is possible.

purchased from, making it impossible to account for intra-sectoral transactions. However, we can add intra-sectoral transactions back into BLS output measures for these comparisons. Using Census microdata, we replicate the value of shipments as closely as possible. Specifically, we calculate plant-level real output as deflated revenues, adjusted for resales and changes in inventories.[10] Thus, we measure plant-level output as:

$$Q_{et} = (TVS_{et} + DF_{et} + DW_{et} - CR_{et})/PISHIP_{it} \qquad (3)$$

where TVS = total value of shipments, $DF_{et} = FIE_{et} - FIB_{et}$ and $DW_{et} = WIE_{et} - WIB_{et}$ are the changes in finished-goods and work-in-process inventories respectively (FIB, FIE = beginning-of-year and end-of-year finished goods inventories and WIB, WIE = beginning-of-year and end-of-year work-in-process inventories), CR = cost of resales, $PISHIP$ = deflator for the value of shipments[11], and the i, e, and t subscripts index industries, establishments and years. Note if either FIE_{et} or FIB_{et} is missing, then we set $DF_{et} = 0$ (DW_{et} is imputed similarly when either of its components are missing). Also, if $Q_{et} \leq 0$, we set:

$$Q_{et} = TVS_{et}/PISHIP_{it} \qquad (3')$$

We measure labor input as total hours worked. For each establishment, the ASM collects the total number of employees and, for production workers, it collects both hours worked and the number of employees. We calculate total annual hours by summing ASM production worker hours and an estimate of nonproduction worker hours, which is calculated as the product of: the number of non-production workers from the ASM, the average weekly hours worked by production workers from the ASM, and the CPS nonproduction/production ratio (similar to equation 1):[12]

---

[10] In practice, subtracting resales does not make much difference because they are only a small fraction of revenue (See Appendix Figure A1 for the differences in output for the manufacturing sector.)

[11] The shipments deflator is constructed as part of NBER-CES data, using price indexes in BEA's GDP-by-Industry data.

[12] Note that it is not necessary to apply the hours-worked-to-hours-paid ratio because ASM hours are hours worked.

$$TH_{et} = PH_{et} + \left((TE_{et} - PW_{et}) \times \frac{PH_{et}}{PW_{et}} \times \left(\frac{\text{AWH}_{\text{NP}}^{\text{CPS}}}{\text{AWH}_{\text{P}}^{\text{CPS}}} \right) \right) \tag{4}$$

where PH = production worker hours, PW = average number of production workers, TE = total

employment, and $\frac{\text{AWH}_{\text{NP}}^{\text{CPS}}}{\text{AWH}_{\text{P}}^{\text{CPS}}}$ = CPS non-production/production average weekly hours ratio. We

calculate establishment-level log labor productivity as:

$$lp_{et} = log(Q_{et}) - log(TH_{et}) \tag{5}$$

2.3. Missing Data and Imputation

As noted above, the ASM microdata are subject to item non-response, and these missing

values are imputed by the Census Bureau. The Census Bureau's imputation methods are designed

to yield accurate published aggregates but do not necessarily preserve the distribution or

adequately reflect the variability of the underlying microdata. There is evidence that certain

imputation methods may affect microdata analyses. However, there are techniques available to

mitigate the effects of imputation on dispersion measures. For example, White, Petrin, and Reiter

(2012) analyze dispersion statistics using classification and regression-tree methods. Foster et al.

(2017a) follow a different approach and address imputation by dropping observations with imputed

data and reweighting the remaining observations. The results from these studies suggest that both

of these approaches reduce within-industry dispersion. For the purposes of this paper, we consider

the entire set of observations in the sample and leave further analysis of these issues for future

work.

3. **Comparison of Micro-Aggregated Data to BLS and NBER-CES Industry Data**

Because one of our goals is to make inferences about the relationship between within-

industry productivity dispersion and the official estimates of average industry productivity, it is

important to know how our micro-aggregated data compare to the corresponding aggregate

measures underlying published official statistics. Based on earlier work comparing similar

business data across the two government agencies, we expect that there will be some systematic

differences between these measures (Elvery et al., 2006). Even though differences in the levels of

the micro and published first moments do not affect our conclusions about dispersion (since we

sweep out industry-year effects), we want to confirm those levels are not too far from each other. If

the levels are close, then it is more reasonable to think of micro-based second moments as

measuring variation around the published BLS first moments.

Below, we compare our estimates to published estimates from the ASM, data from the BLS

productivity program, and data from the NBER-CES database, covering the 1997-2000 period. The

NBER-CES database is used for comparisons in this section and should be thought of as equivalent

to the official published ASM and CM statistics upon which it is based.[13] We start by comparing

input and output measures, and then we compare labor productivity measures.

3.1. Comparing Input and Output Measures

Figure 1 shows the total number of employees in the manufacturing sector from the

different series. The first thing to note is that employment levels based on ASM microdata (using

ASM sample weights) are significantly lower than the published ASM and BLS estimates because

they exclude the "non-mail" stratum—small establishments that are not sampled by the ASM. The

published ASM series includes adjustments for the non-mail stratum and is much closer to the BLS

estimates.

In order to construct indices closer to the published series, we construct an alternative set of

weights. As noted, the weighted sample totals calculated from the ASM (using ASM sample

weights) is by construction not equal to the published total since there are additional adjustments in

---

[13] More information on the NBER-CES Manufacturing Industry Database can be found at
http://www.nber.org/nberces/. The NBER-CES series we use ends in 2010 (accessed June 2016).

the latter for the non-mail cases. Fortunately, there are very small establishments in the ASM

sample each year that are below the thresholds for non-mail cases. This occurs given the volatility

of business size that is especially prevalent for small businesses.[14] This implies that there is

effectively coverage of all size businesses in the ASM sample (e.g., there are ASM establishments

in any given year and industry with 1-4 employees even though this is typically below the ASM

sample threshold). To construct an alternative set of weights we use the LBD. Specifically, we

define the manufacturing universe using the LBD and use LBD data to estimate the probability that

an establishment is included in the ASM sample and use these probabilities to construct inverse

propensity score weights (see the Appendix for a full discussion of the weighting procedures).

These weights (PSW hereafter) yield weighted sample totals consistent with the LBD.[15]

Moreover, as can be seen in Figure 1, the micro-aggregated employment series using PSWs yields

totals that align with the published BLS and ASM more closely than that using ASM weights.[16]

Our approach seeks to replicate the LBD (BR) totals rather than the published ASM.[17]

Despite differences in levels, the BLS-Employees, the ASM Published, and the micro-

aggregated CMP–PSW series all have similar downward trends. The CMP–PSW series lies above

the BLS series. The two series start to converge around 2003-2004, and they lie on top of each

other starting in 2007. The CMP–PSW series also lies above the published ASM series.[18]

---

[14] That is, of the establishments that are selected for the ASM sample based on size in the prior Census and given volatility, many have fallen below the size threshold for sample inclusion in the actual ASM sample years.

[15] In addition, unreported results suggest that PSW does a good job matching the industry/year-specific size and age distributions of the LBD.

[16] We explored the possibility of benchmarking CMP employment (based on the manufacturing universe in the Census Business Register) to BLS employment (based on the manufacturing universe in the BLS Quarterly Census of Employment and Wages). While this benchmarking, by definition, improves the correlation between labor (employment and hours) measures between BLS and CMP, it actually decreases the correlation between BLS and CMP output and measures of other inputs, which are both based on the manufacturing universe in the Census Business Register.

[17] An alternative approach would be to target the published ASM. For this approach, we could use imputed data for the non-mail cases in the ASM. This is already done for the post 2003 data although White, Petrin, and Reiter (2012) raise a variety of questions about Census imputation procedures. Given these concerns, we think a better procedure would be to use an imputation model like that developed for White, Petrin and Reiter (2012) for the non-mail cases. We plan to explore this possibility in future work. This will require imputing data for non-mail cases pre and post 2003.

[18] Although trends are similar, the published ASM series has a spike in employment in 2007.

We next compare total manufacturing hours between the BLS series, the CMP–PSW series, and the NBER series (Figure 2). To increase comparability of the NBER series to the other two series, we estimate nonproduction worker hours using the CPS nonproduction/production worker hours ratio. Again, the series have similar downward trends; however, the CMP hours series lies slightly above the BLS and NBER series. The higher level of the CMP series relative to the BLS series may seem a little surprising because, as noted earlier, the BLS hours measure includes the hours worked by unincorporated self-employed workers and unpaid family workers, who are not included in the CMP hours; but these workers account for only a small fraction of manufacturing employment. [19] Comparing Figures 1 and 2, we can see that much of the difference in total hours between the BLS and CMP series is driven by differences in employment. In Figure 2(b), we see that the growth rates of employee hours exhibit similar trends, except for 2005-2007, when the NBER series diverges.

Figure 3 compares the three real output series. As noted earlier, the ASM collects data on sales but not who products are sold to, which makes it impossible to account for intrasectoral transactions at the establishment-level. Therefore, we add intrasectoral transactions back into the BLS output series for these comparisons. [20] In all three series, we use the BLS implicit price deflator. The BLS and NBER series track each other closely, with the NBER series being slightly higher and having a slightly higher growth rate between 2003 and 2008. The CMP series is much higher than the BLS and NBER-CES series, and the timing of output growth differs somewhat from the other two series.

The top panel of Table 1 shows correlations between the three data sources of hours and output for the total manufacturing sector. Hours and output, both in levels and growth rates, are highly correlated across data series, ranging from 0.92 to 0.99, despite the different sources of data

---

[19] See Figure A2 for the differences in the series with and without the self-employed and unpaid family workers.
[20] Another difference that we do not account for is the inclusion of unincorporated self-employed output in the BLS series.

used to measure labor hours. The bottom half of the table shows the average of the within-industry

correlations for 4-digit NAICS industries, calculated over the 14 years of the sample and weighted

using the share of BLS employees in each industry. Not surprisingly, the correlations are lower

than for total manufacturing, but they are still high—above 0.6—for labor and output, both in

levels and growth rates. Thus, despite differences in data sources and methodologies, we conclude

that micro-aggregated data largely resemble published aggregate data.

3.2. Comparing Labor Productivity

We calculate labor productivity growth as the change in the log-level labor productivity.[21]

Figure 4 shows that labor productivity growth rates for the manufacturing sector are similar in the

three series between 1998 and 2002 and for 2010, but differ from 2003-2009. These differences

largely follow from the differences in output trends we saw in Figure 3. Table 2 shows that the

correlations between the BLS and CMP series are not as strong for labor productivity as for the

underlying output and hours data. This should not be surprising, because differences in the timing

of output growth across series do not coincide with the differences in hours growth across series.

This comparison of inputs, output, and labor productivity serves as an important backdrop

to the statistics on within-industry dispersion we wish to produce. Although there are some

differences between the BLS and the micro-aggregated series, they are close enough to each other

to allow us to make meaningful inferences about the relationship between within-industry

dispersion and BLS's published estimates of industry productivity growth.

4. Productivity Dispersion

For our analysis of productivity dispersion, we focus on levels rather than growth rates.

---

[21] The official BLS productivity series are calculated using percentage changes in the index, and thus the BLS series we refer to here differs from the published series. Additionally, the official total manufacturing productivity series is published by the BLS Division of Major Sector Productivity, whereas the data here are aggregated from industry data provided by the BLS Division of Industry Productivity Statistics.

Because we are interested in comparing within-industry dispersion of productivity across industries and over time, it is necessary to account for industry differences in average productivity. To do this, we normalize productivity across industries by expressing establishment-level productivity relative to the corresponding industry-year averages. Specifically, we calculate establishment-level productivity as a deviation from average productivity in that establishment's 4-digit industry.[22] The interpretation of normalized establishment-specific productivity levels is intuitive: they tell us how far above or below the mean the establishment sits in the productivity distribution.

We use the interquartile range (IQR) as our primary measure of dispersion, because it is intuitive and easy to interpret. The IQR shows how much more productive an establishment at the 75th percentile of the productivity distribution is than an establishment at the 25th percentile. The standard deviation may seem like an obvious alternative to the IQR, but, in addition to being harder to interpret, it is known to be more sensitive to outliers than quantile-based dispersion measures. We also report the 90-10 differential, as well as the 50-10 and 50-90 differentials.

Table 3 summarizes the dispersion of within-industry labor productivity across 4-digit industries.[23] We compute our dispersion measures at the 4-digit industry level in each year, and then compute summary statistics of industry-specific dispersion over the 1,204 industry-year observations. The mean IQR of 0.92 indicates that establishments at the 75th percentile are about 2.5 times as productive as those at the 25th percentile, in the average industry.[24] Further, establishments in the 90th percentile are more than 6 times as productive as those at the 10th percentile. These ranges are somewhat larger than those found by Syverson (2004b)—he found multiples of about 1.9 and about 4.0 using hourly value-added —but they are generally in line with

---

[22] These are weighted averages (using our propensity score weights, see appendix) where establishment-level productivity is expressed as a deviation from average productivity in that establishment's 4-digit industry. This normalizes productivity across industries and thereby accounts for industry differences in average productivity.
[23] We present standard deviations in Table 3 but do not discuss these results.
[24] The 2.5-fold difference is obtained as $e^{0.92} \approx 2.51$.

his results. Differences among establishments tend to be somewhat greater as we move further away from the center of the support of the within-industry productivity distribution, as shown by the higher 50-10 and 90-50 differentials in Table 3. These results suggest that productivity differences between plants remain large as we move away from the center of the support of the within-industry productivity distribution.

Figure 5 summarizes how within-industry dispersion in labor productivity—measured as the IQR—varies across industries and over time. The mean and median IQR are fairly close to each other, but the large differences between the 25th and 75th percentiles indicate that there is a lot of variation in the IQR across industries. For example, in 2002, in the industry at the 75th percentile of the IQR distribution, the productivity difference between establishments at the 75th and 25th percentiles is about 2.9 fold. In the same year, the corresponding multiple was about 2.1 for the industry at the 25th percentile of the IQR distribution. These differences in IQRs suggest that there are factors that differ by industry that generate this "dispersion in dispersion." These factors could include differences in shocks, adjustment costs, distortions, technology, and distributions of capital intensities.

There does not appear to be a discernable upward or downward trend over time in the mean and median IQR or in the 75th and 25th percentiles of the IQR distribution. This result differs from Decker et al. (2018), who find a rising trend in productivity dispersion using data from the Business Register. We can also see from Figure 5 that the volatility of the mean and median IQR over time is smaller than the variation across industries. Still, we do find time series variation that is likely to be important. For example, the median IQR in 1998-1999 is about the same as the 75th percentile IQR in 2003-2006. In addition, the 25th percentile IQR in 1998-1999 is about the same as the median IQR in 2003-2006. It is also worth noting that there is more variation in the IQR of industries that have the greatest dispersion (those in the 75th percentile of IQRs)—although they are not the same industries each year). Looking at variation over the business cycle, we see no

clear pattern.[25] Dispersion increased during the Great Recession, which is consistent with Kehrig

(2015), who finds the cross-sectional dispersion in productivity is countercyclical and argues this

is important for understanding the business cycle. However, dispersion fell during the 2001

recession that followed the dot-com boom.

For the remainder of this section, we consider two extensions to our analysis. We first

examine how our results change when we weight establishments using propensity score weights

and employment weights. This weighting paints a potentially different picture of dispersion

because there may be differences between the dispersion of different size groups..[26] Our second

extension is to examine the tails of the productivity distribution. Because a substantial portion of

wage inequality is driven by the upper tail of the distribution and by increasing between-

establishment wage differentials, one might expect significant dispersion at the upper tail of the

productivity distribution.

4.1. Employment-Weighted Dispersion Measures

Figure 6 replicates the dispersion measures in Figure 5 using employment weights and

propensity score weights.[27] The most notable feature of Figure 6 is that there is significantly less

year-to-year variation in all of the dispersion measures. The median hovers around 0.8, while the

25th percentile hovers around 0.7. Employment-weighted dispersion in the mean industry and at

---

[25] It is important to emphasize that following the movements of percentiles in Figure 5 is informative about industry dynamics only if it is based on a distribution where we can assume that industry-year observations are i.i.d., i.e. compositional changes are not important. If compositional changes are important then such a figure may be misleading, because it confounds genuine and compositional changes. In order to make inference about the business cycle behavior of dispersion, one should focus on genuine changes in dispersion, which is calculated as the average of the growth rates of dispersion measures. We found significant churning in the IQR distribution, as shown by nonzero transition probabilities across quintiles, see Table 6. Therefore, caution is necessary when interpreting Figure 5.
[26] The PSWs are roughly equal to the number of establishments that the observation represents, whereas the employment weights are roughly equal to the number of workers that the observation represents. If all establishments employed the same number of workers, the two measures would generate the same results. To illustrate, suppose that all workers in the same establishment produce the same output per hour. Then the employment-weighted IQR tells us how much more productive a worker at the 75th percentile is compared to a worker at the 25th percentile. Or, more correctly, it tells us how much more productive the establishment that employs workers at the 75th percentile is compared to the establishment that employs workers at the 25th percentile.
[27] The weights are defined as the product of the inverse propensity score weight and the employment weight.

the 75th percentile is generally lower relative to Figure 5 and exhibits a slight upward trend, which indicates that dispersion has increased among large establishments in high-dispersion industries. This upward trend is likely what is driving the increase in the mean IQR. These employment-weighted results are consistent with the findings of Decker et al. (2018), Bils et al. (2017), and Foster et al. (2017a).

4.2. Dispersion in the Tails

Turning to the tails of the productivity distributions, a distinctive feature of the within-industry productivity distribution is that mean and median dispersion in the tails (the 99-to-90 ratio in Figure 7 and the 10-1 ratio in Figure 9) is of about the same magnitude as mean and median dispersion in the center of the support (the 75-to-25 ratio in Figure 5). This is remarkable given that each tail covers only one-fifth as many establishments as the IQR. In addition, we see that dispersion in the right-tail (Figure 7) is less volatile and shows a more positive trend than the IQR (Figure 5). The employment-weighted distributions tell a similar story (Figure 8). The main difference is that mean and median dispersion are lower than the unweighted mean and median dispersion, which suggests that dispersion among larger establishments is generally lower and less volatile.

In contrast to the right tail, the left tail (the 10-1 ratio in Figure 9) is considerably more volatile and does not exhibit an upward trend. Mean dispersion among less productive establishments tends to be somewhat higher than in the middle of the distribution (see Table 3). However, like the IQR, the left tail does not have a positive trend. The greater volatility is not too surprising given establishments at the lower tail of the distribution are more likely to exit, which may imply greater year-to-year changes in the composition of these establishments. The employment-weighted dispersion measures (Figure 10) tell a similar story, although they are slightly less volatile. The employment-weighted mean and median dispersion measures are lower

than the unweighted measures, indicating that dispersion among larger establishments is lower even among the least productive businesses.

These findings highlight the importance of looking at the entire distribution. The IQR is a convenient measure that covers half of the distribution. However, there is just as much dispersion in the upper and lower tails as there is in the middle. We also see that weighting matters. Employment weighting tends to reduce both cross-sectional dispersion and its volatility.

4.3. Dispersion and the Olley Pakes Decomposition

As noted in the Introduction, dispersion in productivity within industries may arise from a number of sources. To provide more perspective on these sources, we use the framework developed by Olley and Pakes (1996) to examine the relationship between dispersion and industry-level productivity. The decomposition of Olley and Pakes applied to labor productivity is based on an index of industry productivity given by the weighted average of the labor productivity of establishments in the industry:

$$P_{it} = \sum_e s_{eit} p_{et} \tag{6}$$

where P_{it} is the industry index in industry i at time t, s_{eit} is establishment e's employment as a share of industry i's employment at time t, and p_{et} is labor productivity for establishment e at time t. This index yields a measure of labor productivity at the industry level that in growth rates is potentially the same conceptually as the official BLS productivity measures. The measures may differ in practice for a variety of reasons, which we have discussed above and will explore in more detail below.

For our purposes, this industry index and the accompanying Olley and Pakes decomposition provide a useful framework to explore some of the properties of productivity dispersion. After a little manipulation and rearranging terms, equation (6) becomes:[28]

---

[28] Olley and Pakes (1996) applied this decomposition originally to TFP.

$$P_{it} = \bar{p}_{it} + \sum_e (s_{eit} - \bar{s}_{it})(p_{eit} - \bar{p}_{it}) \tag{7}$$

The bars over the variables represent the cross-sectional (unweighted) means across all establishments in industry i. Thus, this industry-level productivity index can be decomposed into two components: average plant-level productivity and the covariance between establishment shares of employment and establishment-level productivity. The covariance term has been interpreted as an indicator of how closely size and productivity are related, which in turn may indicate how efficiently resources are allocated; the larger the covariance is, the greater the share of employment in more-productive establishments.

Interpreting the OP covariance term as a measure of allocative efficiency is model dependent and also depends on whether the decomposition is applied to measures of labor productivity or total factor productivity (TFP) (e.g., see Decker et al., 2018). For example, in a frictionless and distortionless framework with specific assumptions about the structure of demand (isoelastic demand) and technology (e.g., Cobb Douglas without overhead labor), dispersion in revenue labor productivity should be zero even when there is dispersion in underlying TFP. This stark finding is consistent with equalization of marginal revenue products under these assumptions. This is the insight of Hsieh and Klenow (2009), who argue that any observed dispersion in such revenue labor productivity measures must reflect some form of distortions or wedges. From this perspective, an increase in distortions or wedges yields an increase in revenue labor productivity dispersion. Bartelsman et al. (2013) show that the OP covariance for labor productivity initially rises with an increase in distortions from the frictionless and distortionless benchmark above but then declines and can become negative. They and Decker et al. (2018) argue that over the empirically relevant range, an increase in frictions and/or distortions leads to a decline in the OP covariance. In other words, over the empirically relevant range they argue that an increase in frictions or distortions would yield an increase in productivity dispersion and a decrease in the OP covariance.

Alternatively, increasing dispersion in labor productivity may be due to an increase in the dispersion of underlying total factor productivity which also yields an increase in the OP covariance (see Decker et al., 2018). The reason is intuitive, as an industry with a higher dispersion of TFP has more scope for industry-level productivity to reflect the extent to which resources are being allocated to more productive businesses. Relatedly, Foster et al. (2017b) find evidence that during periods of rapid innovation in an industry, first entry, then productivity dispersion, and then productivity growth rises after a shakeout period. As they argue, this might lead to first a decline and then an increase in the OP covariance within an industry. The decline might occur initially during a period of intense experimentation but then rise during and subsequent to the shakeout period.

In short, some hypotheses predict a positive relationship between dispersion and the OP covariance while others predict a negative relationship. We explore the empirical relationship between the covariance term and dispersion in our data on 4-digit manufacturing industries. Specifically, we estimate the following regression:

$$OP_{it} = \mu_i + \beta iqr_{it} + \varepsilon_{it} \qquad (8)$$

where OP_{it} denotes the covariance term in industry i in time t, iqr_{it} denotes the interquartile range of plant-level log labor productivity, and μ_j denotes industry-fixed effects. This regression specification is intended to be descriptive, yielding information about conditional correlations.

Table 4 summarizes the results from the regression. The first entry in the column labeled "β_{iqr}", which is based on variation between industries, indicates that a one percentage point increase in dispersion is associated with a 0.21 percentage point increase in the OP covariance term. The second entry in the column indicates that a one percentage point increase in the interquartile range within the average industry is associated with a 0.12 percentage point higher covariance in the average industry. Apart from differences in magnitude, both sets of equations tell us the same story—that dispersion is positively correlated with the covariance term across

industries. These patterns alone are not sufficient to identify precisely the mechanisms at work, but our findings provide support for hypotheses consistent with a positive relationship between dispersion and the OP covariance.

4.4. Industry Productivity Indices: Micro and Macro Based

The OP decomposition can also provide further insight into the relationship between the industry index constructed from micro data (i.e., (6) and (7)) and the official data on industry productivity growth published by the BLS. If the BLS used exactly the same source data for the macro growth indices as that used to produce the micro based index in (6), the micro and macro indices would be equivalent.[29] Even if differences in source data break this equivalence, we can still shed light on the relationship between the micro and macro based industry indices by first-differencing equation (7) and regressing aggregate industry productivity growth rates published by BLS on the growth in $\bar{p}_{it}$ and OP_{it}, which we calculate using plant-level data. This exercise is useful for at least two reasons. First, the coefficients of $\bar{p}_{it}$ and OP_{it} are direct indicators of the correlation between micro-aggregated moments and published labor productivity statistics. Second, it further highlights the connection between the OP covariance term and aggregate productivity dynamics.

Our estimates in Table 5 suggest significant and positive relationships between the components of the micro based index and macro based BLS published industry-level growth statistics. The between-estimator indicates that an industry for which the OP covariance term is one standard deviation (0.19) above the sample mean exhibits 0.14 percentage point higher measured

---

[29] That is, if the source data used to produce the micro productivity estimates and the BLS official data were exactly the same then the productivity index from (6) could be used to generate the growth rate measures BLS publishes. However, in order to avoid index number problems (i.e., to make the OP index and decomposition unit free) the index in (6) is typically implemented as the weighted average of log micro productivity. Using a geometric mean to aggregate to the industry level is different from the implicit arithmetic (sum) methodology used by the BLS methodology. In unreported results, we have found the arithmetic and geometric mean based indices to produce similar results. The implication is that the primary difference in the macro and micro based indices is the source data.

BLS productivity growth. On the other hand, the within-estimator indicates that if the OP term increases by one standard deviation, then BLS measured productivity growth will be approximately 0.07 percentage point higher in that industry.[30] The between-estimator results also imply that an industry with a 1 percentage point higher average plant-level productivity is associated with 0.75 percentage point higher BLS measured productivity growth. The within-estimator implies that a 1 percentage point increase in the average plant-level productivity within an industry yields a 0.32 percentage point increase in BLS measured productivity growth. These results imply that there is a tight relationship between the BLS macro based indices and both components of the OP decomposition of the micro based indices.

4.5.   *A Schematic Description of the Beta-Product*

Our research leads us to propose the following beta-product, which the two agencies could update on an annual basis.[31] The prototype dataset would contain a balanced panel of statistics summarizing the within-industry distributions of plant-level productivity measures such as hourly revenue, hourly value added, and multi-factor productivity.[32] Core statistics about the first moments, standard deviations, interquartile and interdecile ranges of the within-industry distributions of plant level productivity indices, as well as elements of standard decompositions, would be included (Table 8 shows a possible format of the published statistics). All data moments would be frequency-weighted; see sections 2.2 and A.2. In addition, the dataset would include employment-weighted versions of the data moments. The 99-to-90 and 10-to-1 ranges are additional indicators under consideration.

---

[30] The results of the within-estimator on levels are fairly similar ($\hat{\beta}_{OP} = 0.79$). If the OP term increases by one standard deviation (0.11), then the level of productivity will be approximately 0.08 percentage point higher in that industry.

[31] The timeliness of the data depends on the release of establishment- and firm-level information. In non-Census years, the ASM is available in the fall of the subsequent year, while the LBD becomes available in spring of the year after. In Census years, microdata become available later. The productivity dataset can be created approximately a month after the underlying micro data becomes available.

[32] The underlying establishment-level data would also be made available to researchers in secure FSRDCs.

Such a dataset could be used for analyzing productivity dynamics at the plant- and industry level or the entire manufacturing sector. Possible empirical uses of the dataset include, but are not limited to, the empirical exercises of this paper, in which we carry out descriptive analyses of the cross-industry distribution of dispersion (see sections 4.1 and 4.2) and investigate the relationship between industry productivity and various micro-aggregated moments (see sections 4.3 and 4.4). Another example of a relatively straightforward empirical analysis is to explore the cyclical behavior of dispersion, by decomposing observed changes into industry-level growth and compositional changes.

5. Conclusions and Future Work

A growing literature uses micro-level data to examine establishment-level productivity dynamics and finds substantial within-industry productivity dispersion. This paper is part of a larger project that seeks to systematically measure and disseminate information about this dispersion within the context of broader statistical measures currently available.

Thus, we start our paper by comparing inputs and output aggregated from micro-level data to BLS aggregates. We start with labor inputs and output (future versions of this paper will add capital, materials, and energy comparisons). Not surprisingly, we find some differences between BLS industry-level data and micro-aggregated ASM data; however, correlations between BLS and micro-aggregated hours and output range from 0.61 to 0.99.

Using these measures of labor inputs and output, we develop a measure of labor productivity growth and examine some of its properties (future versions of this paper will examine multi-factor productivity growth, and hourly value-added growth). Correlations between BLS and micro-aggregated labor productivity growth are lower than the corresponding hours and output correlations, around 0.54. We find large within-industry dispersion in labor productivity: establishments at the 75th percentile are about 2.3 times as productive as those at the 25th

percentile on average. In addition, we find significant dispersion in that within-industry dispersion across industries. Dispersion in dispersion over time is small by comparison, but it is likely still important. Our results also indicate that average dispersion depends on where we measure it: average dispersion is greater as we move further away from the center of the support of the within-industry productivity distribution. Specifically, average productivity differences across establishments are greatest in the right tail. Similar to what we find for average dispersion, the dynamics of these measures depend on where we measure productivity differences. We find evidence that dispersion among the most productive establishments has been increasing during our sample period, while productivity differences in the left tail do not show these patterns. This suggests that positively trending dispersion found in earlier studies may be a consequence of the dynamics of the most productive establishments. Our analysis suggests these patterns are sensitive to how dispersion is measured. We find that employment weights generally imply smaller, less volatile productivity differences among establishments in the entire distribution. We also find that, on average, weighted dispersion among more productive establishments shows a more pronounced positive trend.

In future work, we plan to explore using information from the LBD on entrants and exiters to produce public-use measures of the within-industry productivity differences between entering and exiting plants (i.e., the net entry component of industry-level productivity). This would give researchers new data to leverage when trying to understand differences over time and across industries in the relative productivity of entrants and exiters.

Contingent on internal and external review, the Census Bureau and the BLS plan to jointly publish public-use measures of within-industry productivity dispersion for industries in the manufacturing sector. A key benefit of making these data available will be to allow researchers without access to the confidential microdata to explore the various possible causes – and effects – of the differences in within-industry dispersion across industries and over time.

References

Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2014. "Demand Fluctuations in the Ready-Mix Concrete Industry," *Journal of Political Economy*, 122(5), 1013–63.

Baily, Martin N., Eric J. Bartlesman, and John Haltiwanger. 2001. "Labor Productivity: Structural Change and Cyclical Dynamics," *The Review of Economics and Statistics* 83(3): 420–433.

Baily, Martin N., Charles Hulten, and David Campbell. 1992. "Productivity Dynamics in Manufacturing Plants," Brookings Papers on Economic Activity. Microeconomics, 187–267.

Barth, Erling, Alex Bryson, James C. Davis, and Richard Freeman. 2016. "It's Where You Work: Increases in Earnings Dispersion across Establishments and Individuals in the United States," Journal of Labor Economics 34(2), S67–S97.

Bils, Mark, Peter J. Klenow, and Cian Rouane. 2017. "Misallocation or Mismeasurement?" unpublished manuscript

Bloom, Nicholas, Fatih Guvenen, David J. Price, Jae Song, and Till von Wachter. 2015. "Firming Up Inequality," NBER Working Paper No. 21199, National Bureau of Economic Research, Inc.

Brown, J. David, Emin Dinlersoz, and John S. Earle. 2016. "Does Higher Productivity Dispersion Imply Greater Misallocation?" Center for Economic Studies Discussion Paper, No. 16–42.

Burdett, Kenneth, and Dale T. Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment," International Economic Review, Department of Economics, University of Pennsylvania and Osaka University Institute of Social and Economic Research Association, 39(2), 257–73.

Davis, Steven J., and John Haltiwanger. 1991. "Wage Dispersion Between and Within U.S. Manufacturing Plants, 1963-1986," NBER Working Paper No. 3722, National Bureau of Economic Research, Inc.

Davis, Steven J., John C. Haltiwanger, and Scott Schuh. 1996. *Job Creation and Destruction*. Cambridge MA: MIT Press.

Decker, Ryan, John Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2018. "Changing Business Dynamism and Productivity: Shocks vs. Responsiveness," NBER Working Paper No. 24236, National Bureau of Economic Research, Inc.

Dunne, Timothy, Lucia Foster, John Haltiwanger, and Kenneth R. Troske. 2004. "Wage and Productivity Dispersion in United States Manufacturing: The Role of Computer Investment," *Journal of Labor Economics*, 22(2), 397–430.

Eldridge, Lucy, Somayah Eltoweissy, Sabrina Pabilonia, and Jay Stewart. 2017. "Estimating Non-production and Supervisory Worker Hours for Productivity Measurement," Unpublished manuscript.

Eldridge, Lucy P., Marilyn E. Manser, and Phyllis F. Otto. 2004. "Alternative Measures of

Supervisory Employee Hours and Productivity Growth," *Monthly Labor Review*, 27(4), 9–28.

Elvery, Joel, Lucia Foster, C.J. Krizan, and David Talan. 2006. "Preliminary Micro Data Results from the Business List Comparison Project," 2006 Proceedings of the American Statistical Associations, Business and Economics Statistics Section [CD-ROM], Alexandria, VA: American Statistical Association.

Fisher, Sylvia, Karen Goldenberg, Eileen O'Brien, Clyde Tucker, and Diane Willimack (2001) "Measuring Employee Hours in Government Surveys," Paper presented to the Federal Economic Statistics Advisory Committee (FESAC), June 7.

Foster, Lucia, Cheryl Grim, and John Haltiwanger. 2016a. "Reallocation in the Great Recession: Cleansing or Not?" *Journal of Labor Economics* 34, no. S1, S293–S331.

Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. 2016b. "Firm-Level Dispersion in Productivity: Is the Devil in the Details?" *American Economic Review* 106(5), 95-98.

Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. 2017a. "Micro and Macro Dynamics of Productivity: From Devilish Details to Insights," NBER Working Paper No. 23666, National Bureau of Economic Research, Inc.

Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. 2017b. "Innovation, Productivity Growth and Productivity Dispersion," prepared for the 2017 CRIW.

Foster, Lucia, John Haltiwanger, and C.J. Krizan. 2001. "Aggregate Productivity Growth: Lessons from Microeconomic Evidence." In *New Developments in Productivity Analysis*, ed. Charles R. Hulten, Edward R. Dean, and Michael J. Harper. Chicago and London: University of Chicago Press, 303–72.

Frazis, Harley and Jay Stewart (2004) "What Can Time Use Data Tell Us About Hours of Work?" Monthly Labor Review 127(12), December, pp. 3–9.

Goldenberg, Karen and Diane Willimack (2003) "Measurement Differences in Key Economic Indicators" JSM Papers and Proceedings.

Griliches, Zvi, and Haim Regev. 1992. "Productivity and Firm Turnover in Israeli Industry: 1979-1988," NBER Working Paper No. 4059, National Bureau of Economic Research, Inc.

Hseih, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics* 124(4), 1403–1448.

Hseih, Chang-Tai, and Peter J. Klenow. 2014. "The Life-Cycle of Manufacturing Plants in India and Mexico," *Quarterly Journal of Economics* 129(3), 1035–1084.

Jarmin, Ron S., and Javier Miranda. 2002. "The Longitudinal Business Database." Center for Economic Studies Discussion Paper, No. 02-17.

Kehrig, Matthias. 2015. "The Cyclical Nature of the Productivity Distribution." Working Paper. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1854401.

Olley, G. Steven and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–97.

Petrin, Amil, T. Kirk White, and Jerome P Reiter. 2011. "The Impact of Plant-level Resource Reallocations and Technical Progress on US Macroeconomic Growth," *Review of Economic Dynamics* 14(1), 3–26.

Syverson, Chad. 2004a. "Market Structure and Productivity: A Concrete Example." *Journal of Political Economy*, 112(6), 1181–222.

Syverson, Chad. 2004b. "Product Substitutability and Productivity Dispersion." *The Review of Economics and Statistics*, 86(2), 534–550.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature*, 49(2), 326–365.

White, T. Kirk, Jerome P. Reiter, and Amil Petrin. 2012. "Plant-level Productivity and Imputation of Missing Data in U.S. Census Manufacturing Data." NBER Working Paper No. 17816, National Bureau of Economic Research, Inc.

Figure 1. Manufacturing Employment Levels, 1997-2010
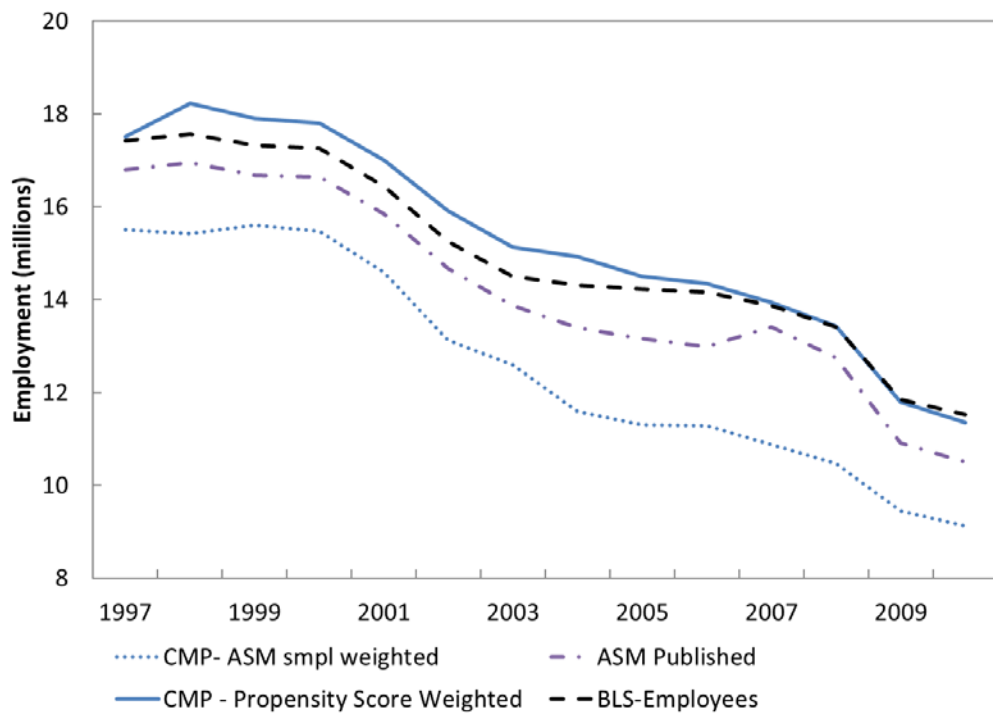
Source: "BLS-Employees" is the annual average of the not seasonally adjusted employment in manufacturing [CEU3000000001, Current Employment Statistics program]. "ASM Published" is the published aggregate employment series from the ASM. CMP denotes micro-aggregated series using the ASM: CMP-ASM Sample Weighted total employment is calculated using ASM sample weights. CMP-Propensity Score Weighted total employment is calculated using our estimated inverse propensity score weights.

Figure 2. Manufacturing Hours Worked, 1997-2010
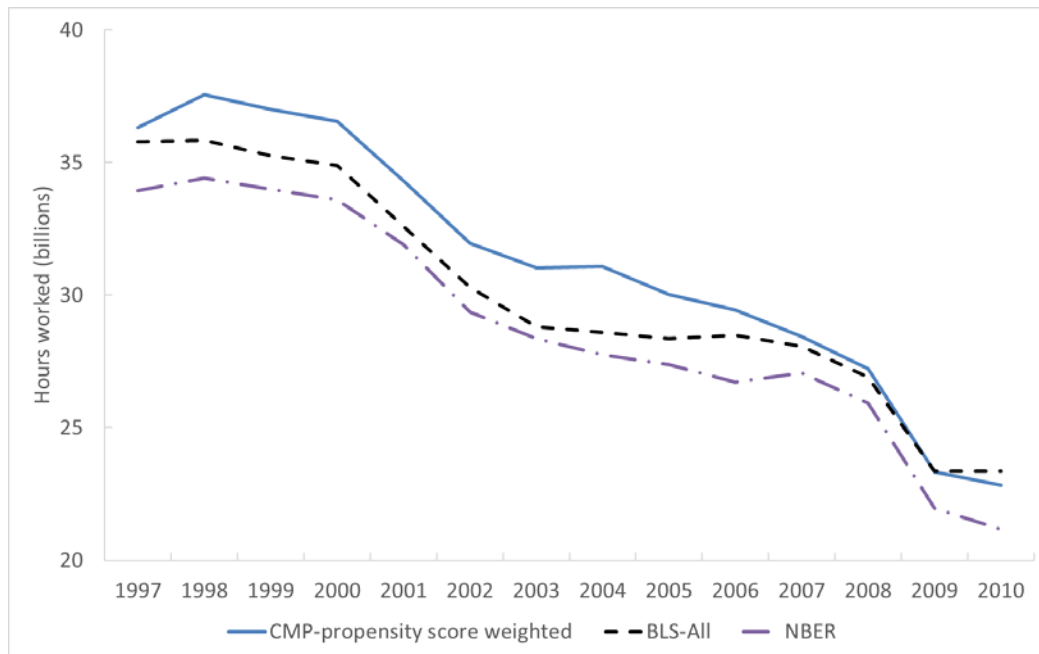
(a)   Levels
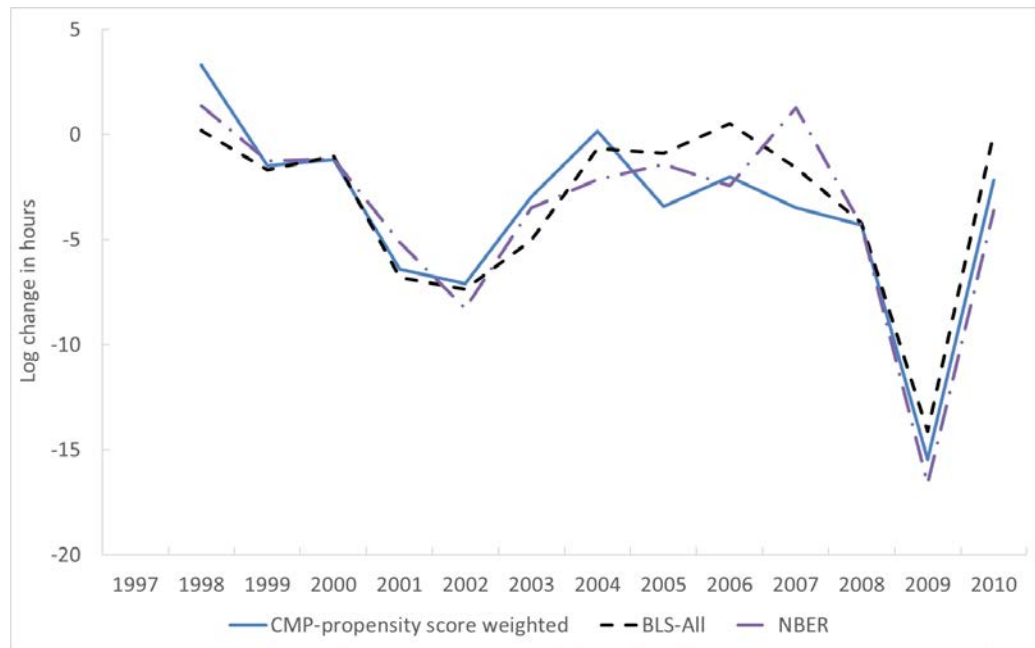
(b) Growth

Source: "BLS-All" denotes total hours, based on authors' calculations from Industry Productivity Program
        data. "CMP – Propensity Score Weight" is the authors' calculations on the ASM. "NBER " is the
        authors' calculations on the NBER database, see Table 7 for details.

**Figure 3.** Manufacturing Output, 1997-2010
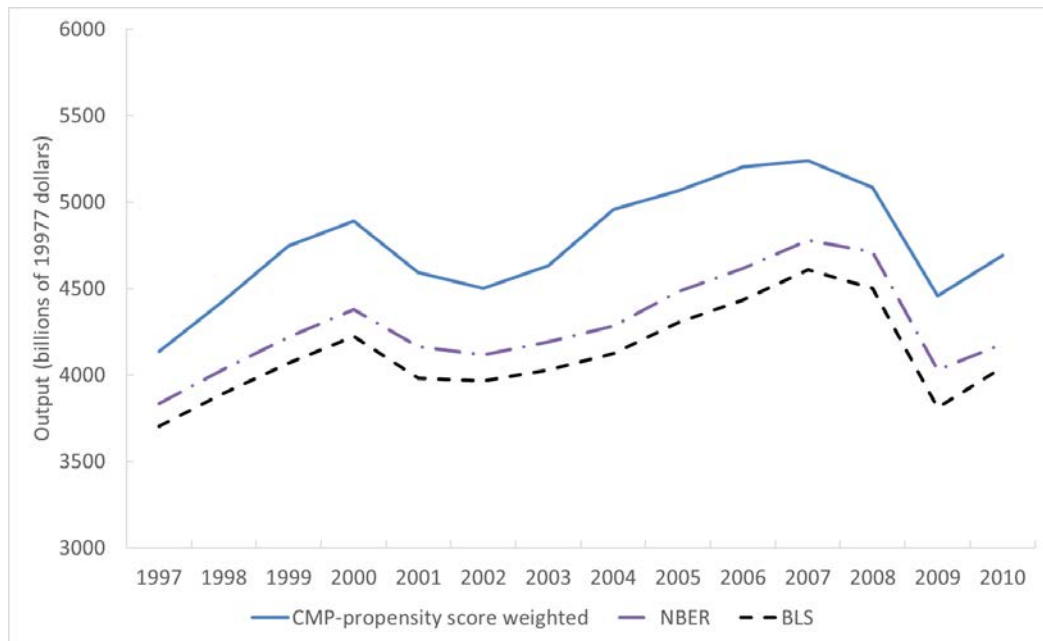

(a) Levels in 1997 dollars
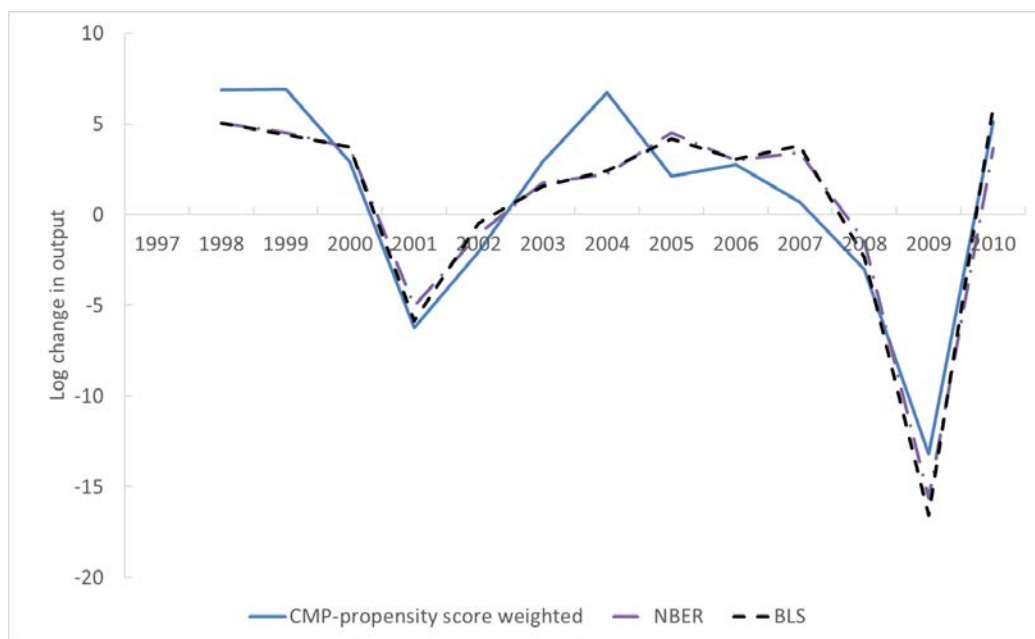
(b) Growth

Source: "BLS" is from the Industry Productivity Program, intrasectoral transactions included. "CMP – Propensity Score Weight" is the authors' calculations on the ASM, cost of resales removed. "NBER" is the NBER Productivity Database, see Table 7 for details.

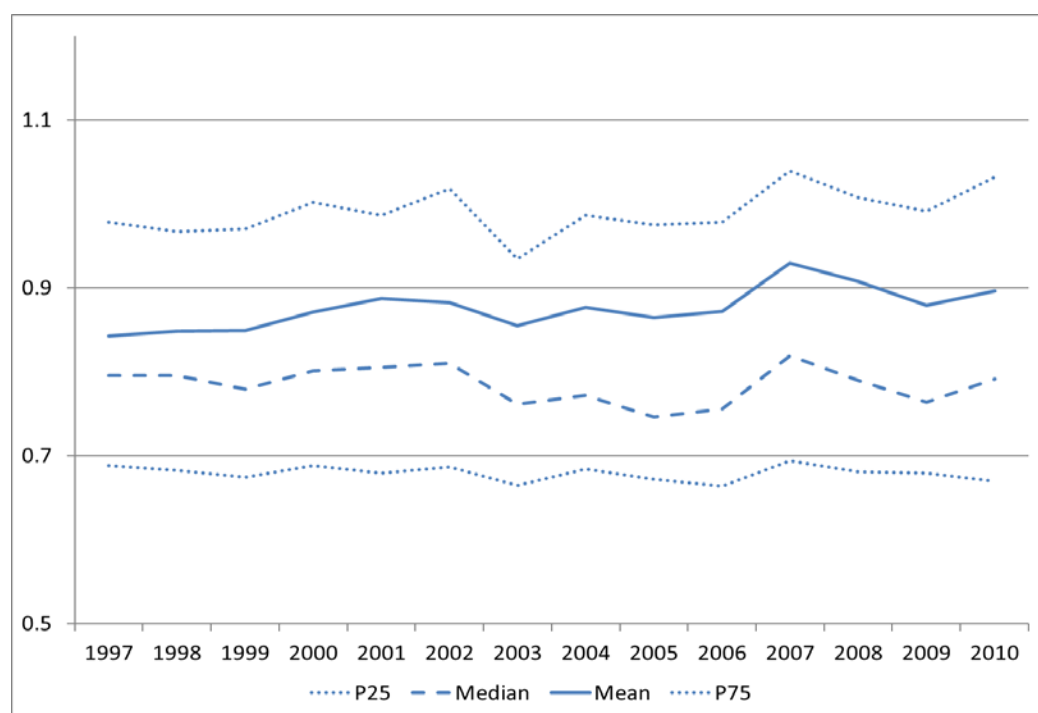**Figure 4.** Growth in Labor Productivity, 1997-2010
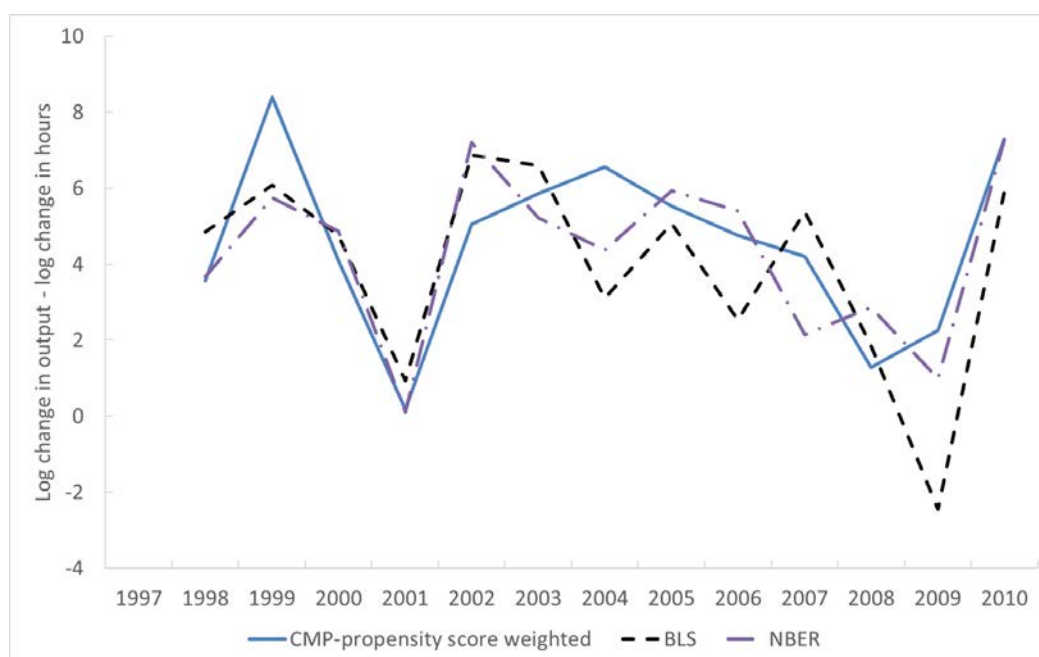
Source: Each graph shows $\Delta lnQ - \Delta lnH$, where the measures of H and Q are taken from figures 2 and 3, respectively. See the notes to Figures 2-3 and Table 7 for details.

Figure 5. Dispersion in Within-Industry IQR of Labor Productivity, 1997-2010
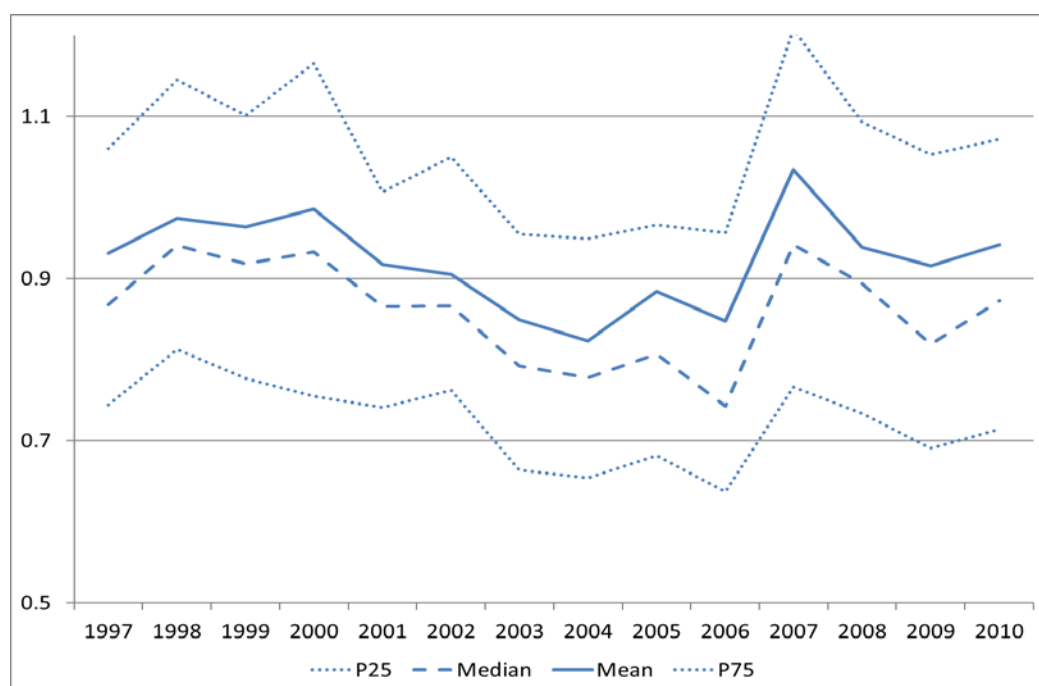

Source: Authors' calculations on the ASM. Notes: The 4-digit NAICS industry mean log labor productivity is subtracted off establishment log labor productivity. Within-industry productivity moments are created at the 4-digit NAICS level, weighted by our propensity score weight. Annual descriptive statistics of industry dispersion are unweighted.

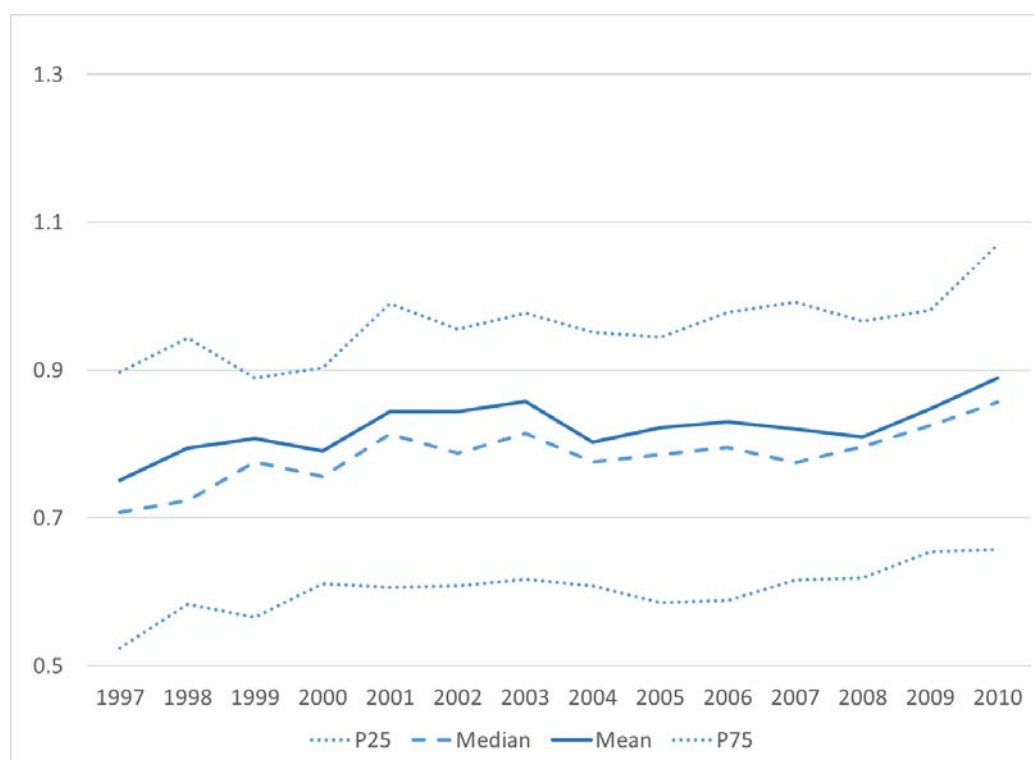**Figure 6.** Dispersion in Within-Industry Employment-Weighted IQR of Labor Productivity, 1997-2010

Notes: See notes to Figure 5.

Figure 7. Dispersion in Within-Industry 99-90 ratio of Labor Productivity, 1997-2010
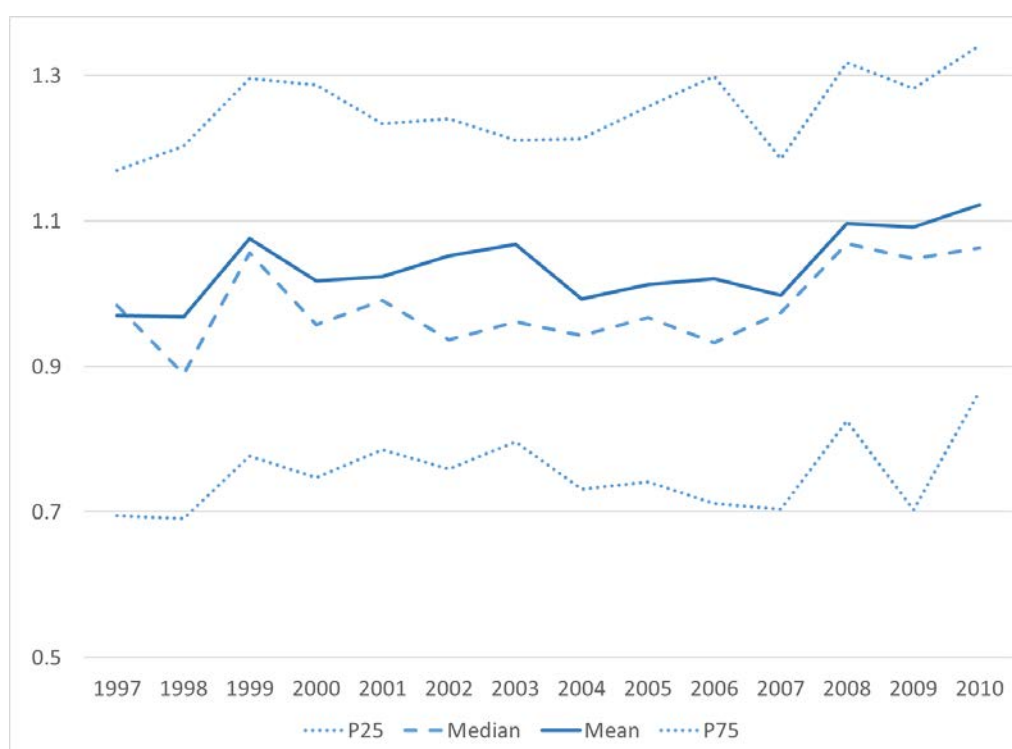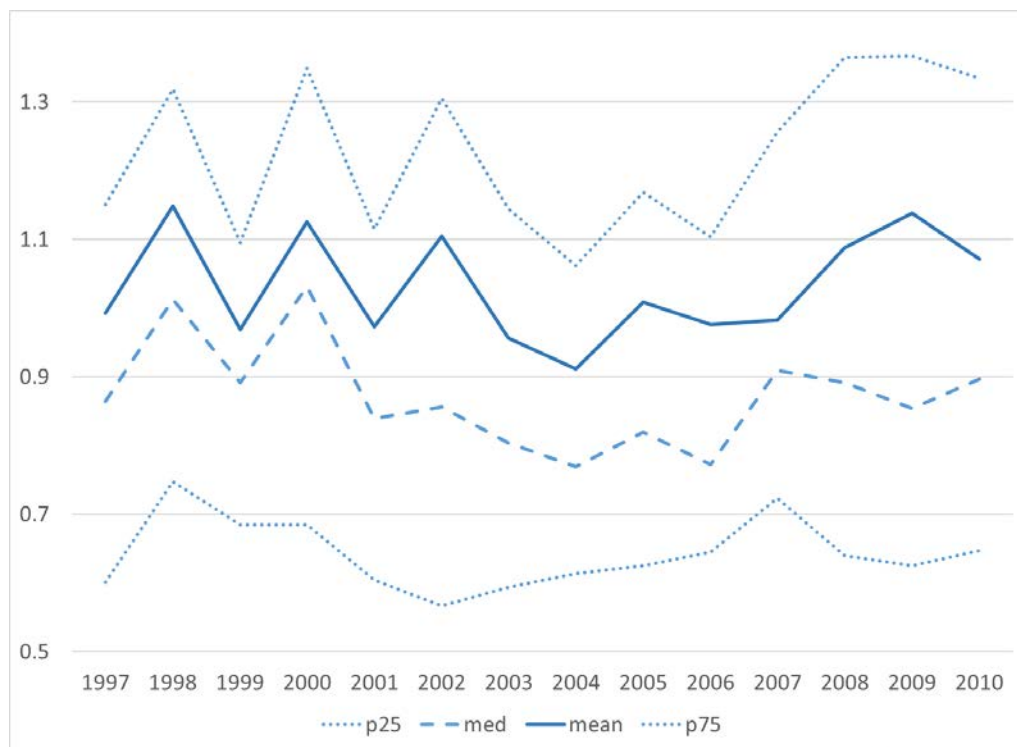

Notes: See notes to Figure 5.

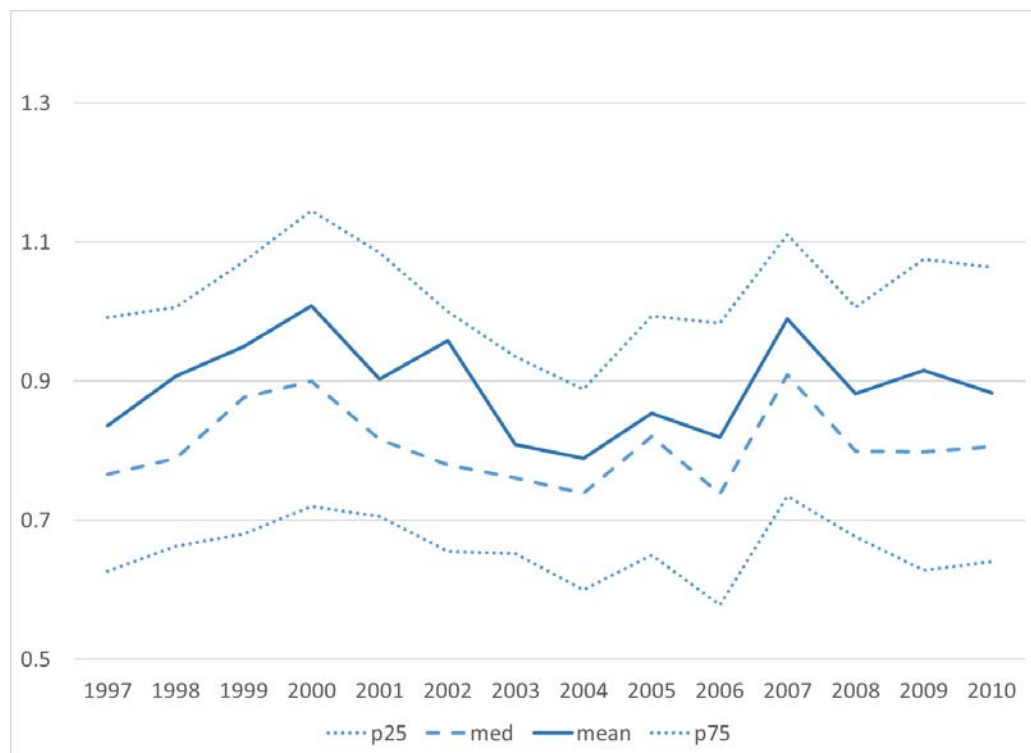**Figure 8.** Dispersion in Within-Industry Employment-Weighted 99-90 ratio of Labor Productivity

Notes: See notes to Figure 5.

Figure 9. Dispersion in Within-Industry 10-1 Ratio of Labor Productivity, 1997-2010

Notes: See notes to Figure 5.

Figure 10. Dispersion in Within-Industry Emloyment-Weighted 10-1 Ratio of Labor Productivity, 1997-2010

Notes: See notes to Figure 5.

Table 1. Hours Worked and Output Correlations between BLS, CMP, and NBER (1997-2010)

	BLS/CMP	BLS/NBER	CMP/NBER
Total Manufacturing			
Hours worked, levels	0.987	0.995	0.992
Hours worked, growth	0.920	0.937	0.920
Output, levels	0.961	0.997	0.961
Output, growth	0.933	0.998	0.929
Average of 4-Digit NAICS			
Hours worked, levels	0.842	0.905	0.903
Hours worked, growth	0.605	0.697	0.717
Output, levels	0.898	0.985	0.894
Output, growth	0.811	0.958	0.779

Source: Authors' calculations on the ASM.
Notes: Resales are removed from output. The BLS output deflator is used, see Table 7 for details.

Table 2. Labor Productivity Growth Correlations between BLS, CMP, and NBER (1997-2010)

	BLS/CMP	BLS/NBER	CMP/NBER
Labor productivity growth, total manufacturing	0.673	0.738	0.765
Labor productivity growth, average of 4-digit NAICS	0.539	0.617	0.673

Source: Authors' calculations on the ASM, see Table 7 for details.

Table 3. Summary of Within-Industry Productivity Distributions (1997-2010)

Within-Industry Productivity Moment	Mean	Standard Deviation	IQR
Labor productivity			
IQR	0.922	0.295	0.332
90-10 Differential	1.907	0.505	0.633
90-50 Differential	1.012	0.328	0.372
50-10 Differential	0.895	0.310	0.341
Standard Deviation	0.807	0.196	0.244

Source: Authors' calculations on the ASM.
Notes: Log labor productivity is calculated as log (output/hours) where hours are BLS-adjusted total hours. The 4-digit NAICS industry mean log LP is subtracted off establishment-level log LP. Within-industry productivity moments are created at the 4-digit NAICS level using propensity score weights. Annual summary statistics of these industry statistics are then created weighting each industry equally. Finally, the numbers shown are means of the annual summary statistic values from 1997-2010 weighting each year equally. Resales have been removed from output, see Table 7 for details.

Table 4. Estimated Effect of Dispersion on the OP Covariance Term

Model	Level	$\widehat{\beta}_{iqr}$	$\widehat{\beta}_{iqr(aw)}$
Between	NAICS6	0.21***	0.11***
Within	NAICS6	0.12***	0.05
Pooled LS	NAICS6	0.14***	0.08***
Between	NAICS4	0.38***	0.32***
Within	NAICS4	0.09	0.07
Pooled LS	NAICS4	0.24***	0.27***

Source: Authors' calculations on the ASM.

Notes: *** Statistically significant at the 1% level. All micro moments are propensity score weighted. $\beta_{iqr(aw)}$ is derived from a regression where iqr_{it} is calculated using the employment-weighted distribution. $OP_{it} = \tilde{p}_{it} - \bar{p}_{it}$, $\tilde{p}_{it} = \sum_{et} \theta_{et} \ln \frac{Q_{et}}{TH_{et}}$, $\bar{p}_{it} = \sum_{et} \ln \frac{Q_{et}}{TH_{et}}$. The last column shows results using weighted dispersion measures.

Table 5. Estimated Effect of $\Delta \bar{p}_{it}$ and Change in the OP Covariance Term on Published ΔP_{it}

Model	N	R^2	$\widehat{\beta}_{\bar{P}}$	$\widehat{\beta}_{OP}$
Between	86	0.73	0.75***	0.82***
Within	1118	0.18	0.32***	0.35***
Pooled LS	1118	0.21	0.36***	0.38***

Source: Authors' calculations on the ASM.

Notes: *** Statistically significant at the 1% level. All micro moments are propensity score weighted. P_{jt} is defined as the log of output per hour (2007=100) as described in the BLS Handbook of Methods (chapter 11, Industry Productivity Measures). $\Delta OP_{it} = OP_{it} - OP_{it-1}$, $OP_{it} = \ln \tilde{p}_{it} - \ln \bar{p}_{it}$, $\tilde{p}_{it} = \sum_{et} \theta_{et} \frac{Q_{et}}{TH_{et}}$, $\bar{p}_{jt} = \sum_{et} \frac{Q_{et}}{TH_{et}}$, where θ_{et} denotes establishment e's share in industry i's total employment.

Table 6. Probability of Transition across Quintiles of the Cross-Industry Distribution of Dispersion (annual averages between 1997 and 2010)

Panel A. Propensity-score-weighted IQR quintiles

	1	2	3	4	5
1	0.56	0.26	0.10	0.04	0.04
2	0.27	0.38	0.23	0.08	0.04
3	0.11	0.22	0.36	0.23	0.08
4	0.05	0.10	0.27	0.34	0.23
5	0.03	0.03	0.09	0.02	0.65

Panel B. Propensity-score- and employment-weighed IQR quintiles

	1	2	3	4	5
1	0.75	0.17	0.06	0.01	0.01
2	0.14	0.60	0.24	0.02	0.00
3	0.03	0.27	0.55	0.16	0.00
4	0.01	0.03	0.16	0.64	0.15
5	0.00	0.00	0.01	0.12	0.87

Source: Authors' calculations on the ASM.

Table 7. Summary of Variables used in Tables and Figures

	Table 1 (Corr.)	Table 2 (Corr.)	Table 3 (Disp.)	Figure 1 (Empl.)	Figure 2 (Hours)	Figure 3 (Output)	Figure 4 (Labor prod)	Figure A1 CR and output	Figure A2 Hours and SE, UFW
				Comparisons					
BLS implicit price deflator used for all estimates	yes	yes				yes	yes		
Shipments deflator used to deflate output			yes						
Cost of resales (CR) removed from CMP	yes	yes	yes			yes	yes	Show difference	
Employees only			yes	yes		N/A		yes	
Include BLS employees and self-employed (SE) and unpaid family workers (UFW) in BLS data only	yes	yes			yes	N/A	yes	yes	Show difference
CPS nonproduction/production hours ratio (even for NBER hours)	yes	yes	yes		yes		yes		yes
BLS intrasectorals included	yes	yes				yes	yes		

Table 8. Structure of Prototype Summary Dataset

t (1)	i (2)	Obs (3)	Industry (NAICS4) (4)	$\bar{p}_{it}$ (5)	$\tilde{p}_{it}$ (θ_{et}) (6)	St.dev (7)	p75-p25 (8)	p90-p10 (9)	St.dev (θ_{et}) (10)	p75-p25 (θ_{et}) (11)	p90-p10 (θ_{et}) (12)	OP_{it} (6)-(5)= (13)
1	1	1	3111									
2	1	2	3111									
⋮	⋮	⋮	⋮									
T	1	21	3111									
⋮	⋮	⋮	⋮									
1	86		3399									
2	86		3399									
⋮	⋮	⋮	⋮									
T	86	1548	3399									

Notes: The dataset would contain information on 86 industries and 18 periods between 1997 and 2014. (N*T = 86*18 = 1548). All micro moments are calculated using the propensity score weighted distribution. $\tilde{p}_{it} = \frac{1}{\sum \theta_{et}} \sum_{et} \theta_{et} p_{et}$, $\bar{p}_{it} = \frac{1}{N_i} \sum_{et} p_{et}$, $OP_{it} = \tilde{p}_{it} - \bar{p}_{it}$, where p_{et} is the log of plant level hourly revenue ($\ln \frac{Q_{et}}{TH_{et}}$) or value added ($\ln \frac{VA_{et}}{TH_{et}}$) or multifactor productivity ($\ln \frac{Q_{et}}{\sum_j^J X_{jet}^{\alpha j}}$), θ_{et} denotes establishment e's share in industry i's total employment.

Appendix

A.1. Properties of ASM samples

The ASM is a 5-year panel of roughly 50,000-70,000 manufacturing establishments and is a sample of establishments drawn from the manufacturing portion of the Census Bureau's Business Register using a probability proportional to size sampling scheme.[33] The largest establishments are sampled with certainty and are included in every panel.[34] Smaller establishments are sampled with a probability less than 1, where the probability increases with establishment size (measured by shipments). The smallest single-unit establishments, which are part of the "non-mail" stratum, are not mailed a form but they are included in the estimates. Due to the desire to reduce reporting burden, the Census Bureau uses administrative records to impute payroll, employment, industry and location from the administrative data for the smallest single-unit establishments, while total value of shipments are imputed using industry averages.[35]

The ASM sample is refreshed every 5 years. New ASM panels are drawn from the Economic Census and begin 2 years after the Census from which it was drawn (years ending in 4 and 9). The sample is also updated annually to include new establishments which are identified on the Census Bureau's Business Register. The Business Register is updated with information from the Economic Census as well as administrative records from the IRS and the Census Bureau's annual Company Organization Survey.

---

[33] More information about the ASM can be found at the Census Bureau's website at http://www.census.gov/manufacturing/asm/.
[34] Prior to 1999, certainty units were establishments with 250 or more employees. In 1999, the cutoff was increased to 500 employees, and in 2004, it was increased again to 1,000 employees. Currently, the 10 largest establishments in an industry are also sampled with certainty. In addition to establishment size, certainty criteria include other characteristics such as industry, cell size, or energy use. For example, Computers, Flat-glass, Sugar, and Small industries (with less than 20 establishments), or establishments with large inventories, assets, fuel/electric expenditures are also sampled with certainty.
[35] Non-mail cases are included in the official estimates and have a weight of one. The survey is designed to tabulate cases from the mail and the non-mail component. The mail component was not designed to estimate the total population.

Data for the ASM are collected in all years except for years ending in 2 and 7, when the ASM data are collected as part of the Economic Census. Data on payroll, employment, industry, and geography for establishments in the non-mail stratum are obtained from administrative records.[36]

The ASM sample is designed to estimate unbiased national level estimates of a skewed population. The large establishments account for over 78% of the total value of shipments and the smallest establishments (5 or fewer employees) account for about 44% of the total number of establishments and only 1% of the total value of shipments. This sample design implies that the establishment counts in various size bins may not reflect those calculated from the LBD.

The ASM sample weights, which are inversely proportional to a shipments-based establishment size measure could, in principle, be used to correct for the effects of the ASM sample design. However, the sample design implies that the weighted sum of shipments from the mail stratum only will not match published totals.[37]

Another important aspect of the sample design is that the composition of establishments changes over time and between sample selections. Any weighting procedure aiming at creating unbiased estimates of productivity dispersion should account for the fact that the sampling probabilities, and therefore the composition of the ASM, change every 5 years. In addition, sampling and non-mail stratum thresholds vary across years.

A.2. Establishment Characteristics and the Probability of Selection into the ASM

The ASM's sample design has important implications for our analysis. For example, the sum of the ASM sample weighted employment or sales might equal total employment or total sales. However, it is not clear that the ASM sample weights are appropriate for our analysis. This

---

[36] Federal regulations require the Census Bureau to limit small establishments' survey response burden.
[37] As mentioned above, only the mail component together with the adjustment for the non-mail stratum yields unbiased estimates of the total population. See Davis, Haltiwanger and Schuh (1996) for more details.

section is devoted to describing our weighting procedure.

To address the effects of the ASM's sample design we construct propensity score weights using the Longitudinal Business Database (LBD). See Jarmin and Miranda (2002) for more details about the LBD. The propensity score weights are constructed from a logistic regression in which we model the relationship between plant characteristics and the probability that an establishment is selected into the ASM. We start by matching establishments in the ASM to LBD establishments by year and "LBD Number."[38] Our dependent variable is a dummy variable that equals one if the establishment is in both the ASM and the LBD for that year and zero if the establishment is only in the LBD. For establishments in the non-mail stratum, the dummy variable is set to 0.

The set of regressors consists of dummy variables that classify each establishment based on its employment and payroll size class, whether the establishment is part of a multi-unit entity, the establishment's industry code, and the interaction between industry and employment size effects. Including industry-size interactions allows us to estimate industry-specific size distributions. These variables are obvious candidates for our logistic regressions because the probability of selection into the ASM sample and the cutoff for the non-mail stratum in the ASM vary by industry and size.

When determining weights, we define industries at the 3-digit NAICS level because the interaction of size indicators and more narrowly-defined industry codes leads to empty cells in smaller industries. Empty-size bins imply that the size distribution cannot be estimated in these industries.[39] When the size distribution cannot be estimated for an industry, propensity scores cannot be calculated because maximum-likelihood estimates of the size effects do not exist. Empty cells can, in principle, be avoided by collapsing size bins, combining similar narrowly-defined industries, or allowing bin definitions to vary across industries. We experimented with the number

---

[38] The LBD Number is an establishment identifier that is consistently defined across both datasets. Although linking the datasets by LBD Number is straightforward, a small percentage of establishment-year observations do not match due to timing issues between the ASM and the LBD.

[39] The size distribution cannot be estimated if all establishments are in the same size bin.

and definition of the size bins and the level of industry aggregation and found that using 3-digit industry codes together with 4 size bins allows us to estimate the size distribution in every industry and year. Allowing for more heterogeneity by using either industry-specific size bins or more narrowly defined industries leads to feasibility problems with the logistic regression.

We defined the size bins so that the resulting distribution allows the lowest size bins to vary over time. That is, in every year and every industry, the 50th percentile of establishments with fewer than 50 employees is used to define bins 1 and 2. For larger establishments, the following bins are defined: 50-99, 100-199, 200+.[40] There are 21 3-digit NAICS industries in the 2002 classification system, which results in 105 industry specific size distributions. We include a continuous size measure in order to allow the weights to vary within these cells. This is necessary to account for possible within-cell compositional changes. Adding 5 payroll classes and 2 groups related to multi-unit status increases the number of cells to 113.[41]

The 2002 change in the industry classification system resulted in missing NAICS-2002 codes for a nontrivial number of establishments in the LBD between 1997 and 2001. For example, the NAICS code is missing if an establishment exited prior to 2002. For these observations, we used imputed NAICS codes.[42] From 2002 on, NAICS codes are available for all establishments in the LBD.

Our inverse propensity score weights generate employment counts that do a good job of matching the trends and cyclical variation on BLS manufacturing employment, but they do not match BLS levels.

---

[40] The payroll size classes are 0-200, 201-500, 501-1000, 1001-5000, 5001+.
[41] If we were to use 4-digit industry, the number of cells would increase significantly. There are 86 4-digit NAICS industries implying 86 different size distributions and 430 industry-size cells. Such an increase in the number of cells yields empty size bins in several industries.
[42] NAICS codes are imputed using a method described in: Teresa Fort. 2013. "NAICS 2002 Code Assignments to the Longitudinal Business Database." Mimeo, Center for Economic Studies.

A.3 Comparison of Hours Measures

In this study, we use hours data from ASM, augmented with data from the CPS. However, for official estimates of productivity growth, BLS uses the CES as its main source of hours data. Although the CES and ASM are establishment surveys, the two surveys differ in what hours data they collect and how they collect it. The best information on these differences comes from studies completed in the early 2000s (Goldenberg and Willimack, 2001; and Fisher et al., 2001). These studies do a nice job of summarizing the differences between the two surveys and how those differences affect estimates of hours worked. In this appendix, we summarize that research and discuss the implications for comparing our estimates to published BLS estimates.

There are some general differences between the two surveys that are worth noting. First, the ASM is an annual survey, whereas the CES is conducted monthly. As a result, the reference periods of the two surveys differ. The reference period for the CES is the pay period that includes the 12th of the month. The CES collects data on the total number of employees, hours for all employees since 2006, the number of production workers, production worker payroll and production worker hours.

In contrast, the ASM has different reference periods for different data elements. For production worker employment, the ASM reference period is the pay period that includes the 12th of the month in the months of March, May, August, and November. These quarterly reports are then averaged into an annual number. The ASM collects employment data for Other Employees only for the pay period that includes March 12th. The implicit assumption is that non-production worker employment does not vary much over the year. Total employment is not collected directly, but rather is equal to the sum of non-production worker employment in March and the annual average of quarterly production worker employment.

Annual total employment in the two surveys can differ if there are seasonal patterns in production worker employment that are missed in the ASM's quarterly reports or if there is a

seasonal pattern to non-production worker employment. We examined this issue using monthly CES data. Specifically, we calculated the average employment for each quarter using CES data, and then calculated the ratio of average quarterly employment to CES employment in the ASM reference months (March, May, August, and November). The ratios were very close to 1, indicating that estimates of average annual employment are the same whether we use four quarterly reports or 12 monthly reports.

There are greater differences in the hours data collected in the two surveys. First, the two surveys use different concepts. The ASM asks employers to report hours *worked*, whereas the CES collects hours *paid*. The main difference is that the CES hours data include holidays, annual leave, and sick leave that were paid but not worked. Thus, we would expect total annual hours reported in the ASM to exceed total annual hours in the CES. For productivity measurement, hours worked is the correct concept, which is why BLS adjusts the CES data using hours-worked-to-hours-paid ratios from the NCS.

The two surveys also differ in how they ask respondents to report hours. The ASM asks respondents to report total annual production worker hours. The CES asks respondents to report total production workers for the pay period that includes the 12th of the month. The hours reports are converted to a weekly number using conversion factors that vary with the number of workdays in the month. Apart from the difference in concept, these two approaches to collecting hours data could result in different estimates of total annual hours. Research by Frazis and Stewart (2004) has shown that people work longer hours during the week that includes the 12th of the month.[43] This would lead to annual hours in the CES being higher than in the ASM and would offset some of the difference due to the difference in concept. Neither survey collects hours data for non-production workers. As noted in the text, nonproduction worker hours are estimated using data from the CPS.

---

[43] Their research examined the accuracy of CPS hours reports by comparing the CPS hours data to hours data from the American Time Use Survey (ATUS). Subsequent research by Eldridge et al. (2017) found differences in some years, but not in others.